

# Chroniques génomiques

## Craig Venter et son exome

Bertrand Jordan



Marseille-Nice Génopole,  
case 901,  
Parc Scientifique de Luminy,  
13288 Marseille Cedex 9, France.  
[brjordan@club-internet.fr](mailto:brjordan@club-internet.fr)

### Le génome, puis l'analyse

Il y a un peu plus d'un an, Craig Venter publiait son « génome diploïde »<sup>1</sup> dans *Plos Biology* [1, 2]. Voici qu'il récidive, cette fois dans *Plos Genetics* [3], avec un article intitulé *Genetic variation in an individual human exome*. Le titre est un peu sibyllin : le terme d'« exome » renvoie certes à l'ensemble des exons (de même que l'on dit génome, transcriptome, protéome...), mais qu'entendent exactement les auteurs par *genetic variation* ? Il s'agit, en fait, des « différences » entre la séquence étudiée et celle déposée au NCBI (*National Center for Biotechnology Information*), laquelle tient actuellement lieu de « séquence officielle » du génome humain. Un titre plus explicite serait *Divergence between an individual's exome and the reference human genome...* En tout état de cause, il s'agit bien une fois encore de l'ADN de Venter, et de son analyse approfondie, ciblée sur les variations trouvées dans l'ensemble des exons par rapport à la « référence NCBI ».

L'accent mis sur cet exome découle de sa liaison, en principe directe, avec le phénotype. Les régions régulatrices de l'ADN, les microARN et autres éléments modulateurs suscitent aujourd'hui beaucoup d'intérêt, mais il n'en reste pas moins que les mutations impliquées dans des maladies génétiques mendéliennes ont quasiment toutes été trouvées dans les régions codantes. Il est donc logique de s'intéresser en priorité à ces 1 % ou

2 % de notre ADN, et l'étude du *J. Craig Venter Institute* est la première à le faire d'une manière globale en prenant en compte l'ensemble des gènes d'une personne. Voyons donc quels sont ses principaux résultats, avant d'aborder les conclusions et de formuler quelques critiques...

### Quatre millions de différences, mille six cents « possiblement significatives »

Prise globalement, la séquence de Venter présente avec celle du NCBI un peu plus de quatre millions de différences, dont trois millions de variations ponctuelles ou *Snip*, le reste correspondant en majorité à des insertions ou délétions (*indel*) [1]. Naturellement, la très grande majorité de ces variations « tombe » en dehors des gènes : le nombre de celles qui se situent dans les exons n'est que de 22 000 environ. Et parmi ces dernières, beaucoup sont silencieuses : la majorité des *Snip* ne changent pas la signification en acides aminés du codon concerné. Restent un peu plus de dix mille *Snip* provoquant un changement d'acide aminé dans la protéine codée, *Snip* que Venter baptise *nsSnip* pour *non-synonymous SNP*. Reste aussi un petit millier d'*indel*, généralement de petite taille. L'article présente alors une analyse détaillée de ces altérations, visant à déterminer si elles sont, ou non, susceptibles de modifier la fonction de la protéine codée par le gène. Pour les *Snip*, l'étude repose sur l'emploi d'algorithmes de prédiction qui indiquent si telle ou telle substitution d'acide aminé modifie la structure ou l'activité de la protéine. Cet examen est complété par la recherche des

<sup>1</sup> « Génome diploïde » désigne ici une séquence déterminée avec une précision suffisante pour distinguer les contributions des chromosomes homologues et donc la part des génomes paternel et maternel.

caractéristiques génétiques de l'allèle considéré : s'il est rare dans la population et présent chez Venter à l'état hétérozygote, cela accroît la probabilité qu'il s'agisse d'une altération délétère et soumise à une sélection négative. Pour les *indel*, on observe une forte prépondérance d'insertions ou délétions affectant un multiple de trois bases, altérations qui respectent la phase de lecture et ont pour effet de raccourcir ou d'allonger la protéine, sans forcément modifier sa fonction. Venter et ses collègues en arrivent finalement à une liste d'environ mille six cents variants « significatifs », dont il est très vraisemblable qu'ils aient des effets nets sur les protéines codées et donc éventuellement sur le phénotype.

### Du génotype au phénotype, une transition délicate

Ainsi, dans leur très grande majorité, les « mutations » (en fait, les divergences par rapport à la référence NCBI) présentes dans l'ADN analysé s'avèrent neutres du point de vue fonctionnel. Même celles qui touchent les protéines et modifient leur séquence en acides aminés sont, à près de 90 %, sans effet sur leur fonction. Restent mille six cent altérations, correspondant souvent à des allèles rares dans la population et/ou présents à l'état hétérozygote chez Venter, susceptibles d'influer sur le phénotype de cet individu. Que peut-on conclure à leur sujet en consultant le corpus considérable amassé par la génétique médicale depuis les années 1980 ? Fort peu de chose, en vérité. En interrogeant OMIM (*On-line Mendelian Inheritance in Man*), le catalogue le plus complet des maladies génétiques humaines et des mutations qui leur sont associées<sup>2</sup>, il s'avère que 682 des *nsSnip* de Venter (parmi le total des 10 000 *nsSnip* répertoriés) tombent dans 443 « gènes de maladies »<sup>3</sup>. Mais seul sept d'entre eux (tous à l'état hétérozygote) correspondent à des mutations déjà répertoriées et donc à un corrélat physiologique connu, constitué en l'espèce par un risque légèrement accru de boulimie, de cancer de la prostate, ou un risque réduit de myélome multiple... Les effets en cause sont faibles (risque relatif compris entre 0,7 et 1,5), et il n'est même pas évident qu'ils s'appliquent à l'individu concerné compte tenu du rôle d'autres gènes et, bien sûr, de l'environnement. Les autres mutations repérées ne correspondent à rien de connu du point de vue de la génétique médicale, et même pour les 81 qui (d'après les analyses de structure) sont censées modifier les propriétés de la protéine, aucune conclusion ne peut en être tirée. Comme le dit l'article « Nous en sommes seulement au début de la mise en correspondance du génotype avec le phénotype, même pour les gènes de maladies bien caractérisés »<sup>4</sup>.

<sup>2</sup> Œuvre incontournable construite depuis les années 1960 par Victor McKusick, récemment disparu.

<sup>3</sup> C'est-à-dire gènes dont les versions altérées peuvent causer ou favoriser la survenue d'une maladie.

<sup>4</sup> « [...] we are only at the beginning of relating genotypes to phenotypes, even for the well-characterized disease genes ».

Au vu des maigres résultats de ce travail, qui a pourtant examiné de manière exhaustive la séquence complète d'un individu et a tenté d'en tirer le maximum d'informations fonctionnelles, on ne peut qu'être très sceptique sur ce que peuvent apporter les profils génétiques à mille dollars proposés par *Navigenics*, *Decode Genetics* ou *23andme* [4]. Cela d'autant plus qu'un examen très attentif des données de séquence (obtenues par la technique classique de Sanger) s'est ici avéré nécessaire pour éviter autant que possible les erreurs : cela a été jusqu'à l'examen manuel des « traces » sorties des séquenceurs *Applied Biosystems* par trois observateurs différents, suivi d'un vote à la majorité - qui a éliminé un bon tiers des *Snip* primitivement repérés. On doute qu'une entreprise cherchant à minimiser ses coûts s'entoure d'autant de précautions...

### Une étape importante

Je n'ai ici présenté que quelques aspects d'un article très riche et qui mérite une lecture attentive. Malgré quelques aspects irritants, comme l'omniprésence du principal intéressé<sup>5</sup> ou l'insertion de quelques phrases dont le contenu d'information est nul (*We also find that within an individual, the basic principles of genetics are followed*), ce travail apporte des informations précises et détaillées, et montre à la fois tout ce que nous apprend l'analyse comparée de deux exomes, et tout ce qu'il nous reste à découvrir, à confirmer, à quantifier avant de pouvoir prétendre déduire d'une séquence d'ADN un phénotype. L'imprudent<sup>6</sup> qui affirmait dans les années 1980 « Donnez moi une séquence et je vous calculerai un individu »<sup>7</sup> a encore du souci à se faire... ♦

### Craig Venter's exome

### RÉFÉRENCES

1. Levy S, Sutton G, Ng PC, et al. The diploid genome sequence of an individual human. *PLoS Biol* 2007 ; 5 : e254.
2. Jordan B. Chroniques génomiques. Les révélations du « génome diploïde » de Craig Venter. *Med Sci (Paris)* 2007 ; 23 : 875-6.
3. Ng PC, Levy S, Huang J, et al. Genetic variation in an individual human exome. *PLoS Genet* 2008 4 : e1000160.
4. Jordan B. Chroniques génomiques. « Génome personnel », gadget ou révolution ? *Med Sci (Paris)* 2008 ; 24 : 91-4.

<sup>5</sup> La séquence de Craig Venter est publiée par cette même personne sous l'égide de l'institut qui porte son nom... Les temps de l'anonymat scientifique et de la pudeur académique sont bien loin !

<sup>6</sup> Dont je n'ai pas réussi à retrouver la trace...

<sup>7</sup> « Give me a DNA sequence, and I will compute a man ».

### TIRÉS À PART

B. Jordan